(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 April 2009 (02.04.2009)

PCT

(10) International Publication Number
WO 2009/039876 A1

(51) International Patent Classification:
G06K 9/68 (2006.01)

(21) International Application Number:
PCT/EP2007/009763

(22) International Filing Date:
12 November 2007 (12.11.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/861,854 26 September 2007 (26.09.2007) US

(71) Applicant (*for all designated States except US*): **FOTO-NATION VISION LIMITED** [IE/IE]; Galway Business Park, Dangan, Galway (IE).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **CORCORAN, Peter** [IE/IE]; Cregg, Claregalway, Galway (IE). **PETRESCU, Stefan, Mirel** [RO/RO]; Bl.73, sc.3, et.2, ap.37, Str. Cap, Gheorghe Ion, nr.4, Sector 4, Mun.Bucuresti (RO). **GANGEA, Mihnea** [RO/RO]; Sos. Nicolae Titulescu Nr. 10, Bloc 20, Scara A et 9, Ap. 44, Sector 1, Bucuresti (RO). **BIGIOI, Petronel** [RO/IE]; 57 Sceilg Ard, Headford Road, County Galway (IE). **POSOSIN, Alexei** [MD/IE]; 28 Arus na Tra, Cois Cuain, Dr. Colohan Road, Galway (IE).

(74) Agents: **BOYCE, Conor** et al.; F.R. Kelly & Co., 27 Clyde Road, Ballsbridge, Dublin 4 (IE).

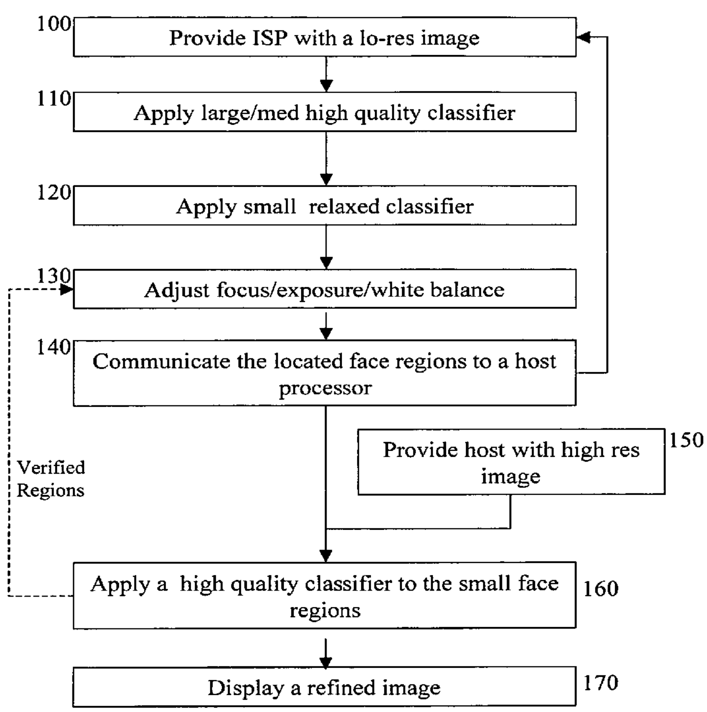
(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH,

[Continued on next page]

(54) Title: FACE TRACKING IN A CAMERA PROCESSOR

Figure 2



(57) Abstract: A method operable in a digital image acquisition system having no photographic film is provided. The method comprises receiving a relatively low resolution image of a scene from an image stream, wherein the scene potentially includes one or more faces. At least one high quality face classifier is applied to the image to identify relatively large and medium sized face regions and at least one relaxed face classifier is applied to the image to identify relatively small sized face regions. A relatively high resolution image of nominally the same scene is received and at least one high quality face classifier is applied to the identified small sized face regions in the higher resolution version of said image.

WO 2009/039876 A1

GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

Face Tracking in a Camera Processor

FIELD OF THE INVENTION

The present invention provides an improved method and apparatus for image processing in acquisition devices. In particular, the invention provides improved face tracking in a digital image acquisition device, such as a camera phone.

BACKGROUND

Figure 1 illustrates digital image acquisition apparatus, for example a camera phone. The apparatus 10 comprises an Image Signal Processor, ISP, 14, which is in general, a general purpose CPU with relatively limited processing power. Typically, the ISP 14 is a dedicated chip or chip-set with a sensor interface 20 having dedicated hardware units that facilitate image processing including image pipeline 22. Images acquired by an imaging sensor 16 are provided to the ISP 14 through the sensor interface 20.

The apparatus further comprises a relatively powerful host processor 12, for example, an ARM9, which is arranged to receive an image stream from the ISP 14.

The apparatus 10 is equipped with a display 18, such as an LCD, for displaying preview images, as well as any main image acquired by the apparatus. Preview images are generated automatically once the apparatus is switched on or only in a pre-capture mode in response to half pressing a shutter button. A main image is typically acquired by fully depressing the shutter button.

Conventionally, high level image processing, such as face tracking, is run on the host processor 12 which provides feedback to the pipeline 22 of the ISP 14. The ISP 14 then renders, adjusts and processes subsequent image(s) in the image stream based on the feedback provided by the host processor 12, typically through an I2C interface 24. Thus,

acquisition parameters of the subsequent image in the stream may be adjusted such that the image displayed to the user is enhanced.

Such acquisition parameters include focus, exposure and white balance.

Focus determines distinctness or clarity of an image or relevant portion of an image and is dependent on a focal length of a lens and a capture area of the imaging sensor 16. Methods of determining whether an image is in-focus are well known in the art. For example, if a face region is detected in an image, then given that most faces are approximately the same size and the size of the face within an acquired image, an appropriate focal length can be chosen for a subsequent image to ensure the face will appear in focus in the image. Other methods can be based on the overall level of sharpness of an image or portion of an image, for example, as indicated by the values of high frequency DCT coefficients in the image. When these are highest in the image or a region of interest, say a face region, the image can be assumed to be in-focus. Thus, by adjusting the focal length of the lens to maximize sharpness, the focus of an image may be enhanced.

Exposure of an image relates to an amount of light falling on the imaging sensor 16 during acquisition of an image. Thus an under-exposed image appears quite dark and has an overall low luminance level, whereas an overexposed image appears quite bright and has an overall high luminance level. Shutter speed and lens aperture affect the exposure of an image and can therefore be adjusted to improve image quality and the processing of an image. For example, it is well known that face detection and recognition are sensitive to over or under exposure of an image and so exposure can be adjusted to optimize the detection of faces within an image stream.

Due to the fact that most light sources are not 100% pure white, objects illuminated by a light source will be

subjected to a colour cast. For example, a halogen light source illuminating a white object will cause the object to appear yellow. In order for a digital image acquisition apparatus to compensate for the colour cast, i.e. perform white balance, it requires a white reference point. Thus, by identifying a point in an image that should be white, for example the sclera of an eye, all other colours in the image may be compensated accordingly. This compensation information may then be utilised to determine the type of illumination under which an image should be acquired.

While adjusting acquisition parameters such as those described above is useful and can improve image quality and processing, the feedback loop to the ISP 14 is relatively slow, thereby causing delays in providing the ISP 14 with the relevant information to rectify the focus, exposure and white balance of an image. This can mean that in a fast changing scene, adjustment indications provided by the host processor 12 may be inappropriate when they are made by the ISP 14 to subsequent images of the stream. Furthermore, typically most of the processing power available to the host processor 12 is required to run the face tracker application, leaving minimal processing power available for carrying out value added processing.

It is an object of the present invention to mitigate the disadvantages associated with the prior art and to provide an improved method of face tracking in a digital image acquisition device.

DESCRIPTION OF THE INVENTION

The present invention provides a method operable in a digital image acquisition system as claimed in claim 1.

BRIEF DESCRIPTION OF DRAWINGS

Embodiments of the invention will now be described, by way of example, with reference to the accompanying drawings, in which:

Figure 1 is a block diagram of a digital image acquisition apparatus in which a preferred embodiment of the present invention may be implemented; and

Figure 2 is a workflow illustrating a preferred embodiment of the present invention.

DESCRIPTION OF PREFERRED EMBODIMENTS

Face tracking for digital image acquisition devices include methods of marking human faces in a series of images such as a video stream or a camera preview. Face tracking can be used to indicate to a photographer, locations of faces in an image or to allow post processing of the images based on knowledge of the locations of the faces. Also, face tracker applications can be used in adaptive adjustment of acquisition parameters of an image, such as, focus, exposure and white balance, based on face information in order to produce improved the quality of acquired images.

In general, face tracking systems employ two principle modules: (i) a detection module for locating new candidate face regions in an acquired image or a sequence of images; and (ii) a tracking module for confirming face regions.

A well-known method of fast-face detection is disclosed in US 2002/0102024, hereinafter Viola-Jones. In Viola-Jones, a chain (cascade) of 32 classifiers based on rectangular (and increasingly refined) Haar features are used with an integral image, derived from an acquired image, by applying the classifiers to a sub-window within the integral image. For a complete analysis of an acquired image, this sub-window is shifted incrementally across the integral image until the entire image has been covered.

In addition to moving the sub-window across the entire integral image, the sub window is also scaled up/down to cover the possible range of face sizes. It will therefore be seen that the resolution of the integral image is determined by the smallest sized classifier sub-window, i.e. the smallest size face to be detected, as larger sized sub-windows can use intermediate points within the integral image for their calculations.

A number of variants of the original Viola-Jones algorithm are known in the literature, such as disclosed in International Patent Application No. PCT/EP2007/005330 (FN143).

In the present embodiment, a face tracking process runs on the ISP 14 as opposed to the host processor 12. Thus, more processing power of the host processor is available for further value added applications, such as face recognition. Furthermore, parameters of an acquired image, such as focus, exposure and white balance, can be adaptively adjusted more efficiently by the ISP 14.

As will be appreciated, face tracking applications carried out on high resolution images will generally achieve more accurate results than on relatively lower resolution images. Furthermore, tracking relatively small size faces within an image generally requires proportionally more processing than for larger faces.

The processing power of the ISP 14 is of course limited, and so the arrangement of face tracking application according to the present invention is optimized to run efficiently on the ISP 14.

In the preferred embodiment, a typical input frame resolution is 160 by 120, and face sizes are categorised as small, medium or large. Medium sized and large sized faces in an image are detected by applying 14x14 and 22x22 high quality classifiers respectively, e.g. relatively long

cascade classifiers or classifiers with a relatively high threshold for accepting a face.

The distance of a subject face from the acquisition apparatus determines a size of the subject face in an image. Clearly, a first subject face located at a greater distance from the acquisition device than a second subject face will appear smaller. Smaller sized faces comprise fewer pixels and thus less information may be derived from the face. As such, detection of smaller sized faces is inherently less reliable even given the proportionally more processing required than for larger faces.

In the preferred embodiment, small sized faces are detected with a relaxed 7x7 classifier, e.g. a short-cascade classifier or classifier with a lower threshold for accepting a face. Using a more relaxed classifier reduces the processing power which would otherwise be required to detect small sized faces.

Nonetheless, it is appreciated that the application of such a relaxed classifier results in a larger number of false positives, i.e. non-face regions being classified as faces. As such, the adjustment of image acquisition parameters is applied differently in response to detection of small faces and the further processing of images is different for small faces than medium or large faces as explained below in more detail.

Figure 2 shows a workflow illustrating a preferred embodiment of the present invention.

On activation, the apparatus 10 automatically captures and stores a series of images at close intervals so that sequential images are nominally of the same scene. Such a series of images may include a series of preview images, post-view images, or a main acquired image.

In preview mode, the imaging sensor 16 provides the ISP 14 with a low resolution image e.g. 160 by 120 from an image stream, step 100.

The ISP 14 applies at least one high quality classifier cascade to the image to detect large and medium sized faces, step 110. Preferably, both 14x14 and 22x22 face classifier cascades are applied to the image.

The ISP 14 also applies at least one relaxed face classifier to the image to detect small faces, step 120. Preferably, a 7x7 face classifier is applied to the image.

Based on knowledge of the faces retrieved from the classifiers, image acquisition parameters for a subsequent image in the stream may be adjusted to enhance the image provided to the display 18 and/or to improve processing of the image. In the preferred embodiment, knowledge of the faces retrieved from the classifiers is utilised to adjust one or more of focus, exposure and/or white balance of a next image in the image stream, step 130.

Knowledge of the faces received from the classifiers comprises information relating to the location of the faces, the size of the faces and the probability of the identified face actually being a face. International Patent Application No. PCT/EP2007/006540 (FN182/FN232/FN214) discusses determining a confidence level indicating the probability of a face existing at the given location. This information may be utilised to determine a weighting for each face to thereby facilitate the adjustment of the acquisition parameters.

In general, a large face will comprise more information than a relatively smaller face. However, if the larger face has a greater probability of being falsely identified as a face, and/or is positioned at non-central position of the image, it could be allocated a lower weighting even than that of a relatively smaller face, positioned at a centre of the image and comprising a lower probability of being a false

positive. Thus, the information derived from the smaller face could be used to adjust the acquisition parameters in preference to the information derived from the large face.

In the embodiment, where only small sized faces are detected in the image, knowledge of the small faces is utilised only to adjust exposure of the next image in the stream. It will be appreciated that although the relaxed classifier passes some false positives, these do not severely adversely influence the adjustment of the exposure.

Focus adjustment is not performed on the next image based on small faces, due to the fact that a lens of the apparatus will be focused at infinity for small faces and there is little to be gained from such adjustment. White balance is not adjusted for small faces because they are considered too small to retrieve any significant white balance information. Nonetheless, each of focus and white balance can be usefully adjusted based on detection of medium and large sized faces.

In the preferred embodiment, once a user acquires a full-sized main image, e.g. by clicking the shutter, and this is communicated to the host, step 150, the detected/tracked face regions are also communicated to the host processor 12, step 140.

In alternative embodiments full-sized images may be acquired occasionally without user intervention either at regular intervals (e.g. every 30 preview frames, or every 3 seconds), or responsive to an analysis of the preview image stream - for example where only smaller faces are detected it may be desirable to occasionally re-confirm the information deduced from such images.

After acquisition of a full-sized main image the host processor 12 retests the face regions identified by the relaxed small face classifier on the larger (higher resolution) main image, typically having a resolution of

320x240, or 640x480, with a high quality classifier, step 160. This verification mitigates or eliminates false positives passed by the relaxed face classifier on the lower resolution image. Since the retesting phase is carried out on a higher resolution version of the image, the small sized faces comprise more information and are thereby detectable by larger window size classifiers. In this embodiment, both 14x14 and 22x22 face classifiers are employed for verification.

Based on the verification, the main image can be adjusted for example, by adjusting the luminance values of the image to more properly illuminate a face or by adjusting the white balance of the image. Other corrections such as red-eye correction or blur correction are also improved with improved face detection.

In any case, the user is then presented with a refined image on the display 18, enhancing the user experience, step 170.

The verification phase requires minimal computation, allowing the processing power of the host processor 12 to be utilised for further value added applications, for example, face recognition applications, real time blink detection and prevention, smile detection, and special real time face effects such as morphing.

In the preferred embodiment, a list of verified face locations is provided back to the ISP 14, indicated by the dashed line, and this information can be utilised to improve face tracking or image acquisition parameters within the ISP 14.

In an alternative embodiment, the verification phase can be carried out on the ISP 14 as although verification is carried out on a higher resolution image, the classifiers need not be applied to the whole image, and as such little processing power is required.

The invention is not limited to the embodiment(s) described herein but can be amended or modified without departing from the scope of the present invention.

Claims:

1. A method operable in a digital image acquisition system having no photographic film, said method comprising:

a) receiving a relatively low resolution image of a scene from an image stream, said scene potentially including one or more faces;

b) applying at least one high quality face classifier to said image to identify relatively large and medium sized face regions;

c) applying at least one relaxed face classifier to said image to identify relatively small sized face regions;

d) receiving a relatively high resolution image of nominally the same scene; and

e) applying at least one high quality face classifier to said identified small sized face regions in said higher resolution version of said image.

2. The method according to claim 1 comprising performing steps a) to c) on a first processor and performing steps d) and e) on a separate second processor.

3. The method of claim 1 wherein each of steps b) and c) include providing information including face size, face location, and an indication of a probability of said image including a face at or in the vicinity of said face region.

4. The method of claim 3 further comprising generating a weighting based on said information.

5. The method according to claim 3 comprising adjusting image acquisition parameters of a subsequent image in said image stream based on said information.

6. The method according to claim 5 wherein said adjusted image acquisition parameters include at least one of focus, exposure and white balance.

7. The method according to claim 5 wherein said subsequent image in said stream is a preview image.

8. The method according to claim 5 wherein said subsequent image in said stream is a main acquired image.

9. The method according to claim 5 further comprising displaying said subsequent image to a user.

10. The method according to claim 2 comprising: further comprising performing value added applications on said high resolution image on said separate second processor.

11. The method according to claim 1 in which said at least one high quality face classifier comprises a relatively long cascade classifier or a classifier with a relatively high threshold for accepting a face and wherein said relaxed classifier comprises a relatively short cascade classifier or a classifier with a relatively low threshold for accepting a face.

12. A digital image acquisition apparatus comprising, a first processor operably connected to an imaging sensor, and a second processor operably connected to said first processor, said first processor being arranged to provide an acquired image to said second processor and said second processor being arranged to store said image,

said first processor being arranged to apply at least one high quality face classifier to a relatively low resolution image of a scene from an image stream, said scene

potentially including one or more faces, to identify relatively large and medium sized face regions, and to apply at least one relaxed face classifier to said image to identify relatively small sized face regions; and

said second processor being arranged to receive a relatively high resolution image of nominally the same scene and to apply at least one high quality face classifier to said identified small sized face regions in said higher resolution version of said image.

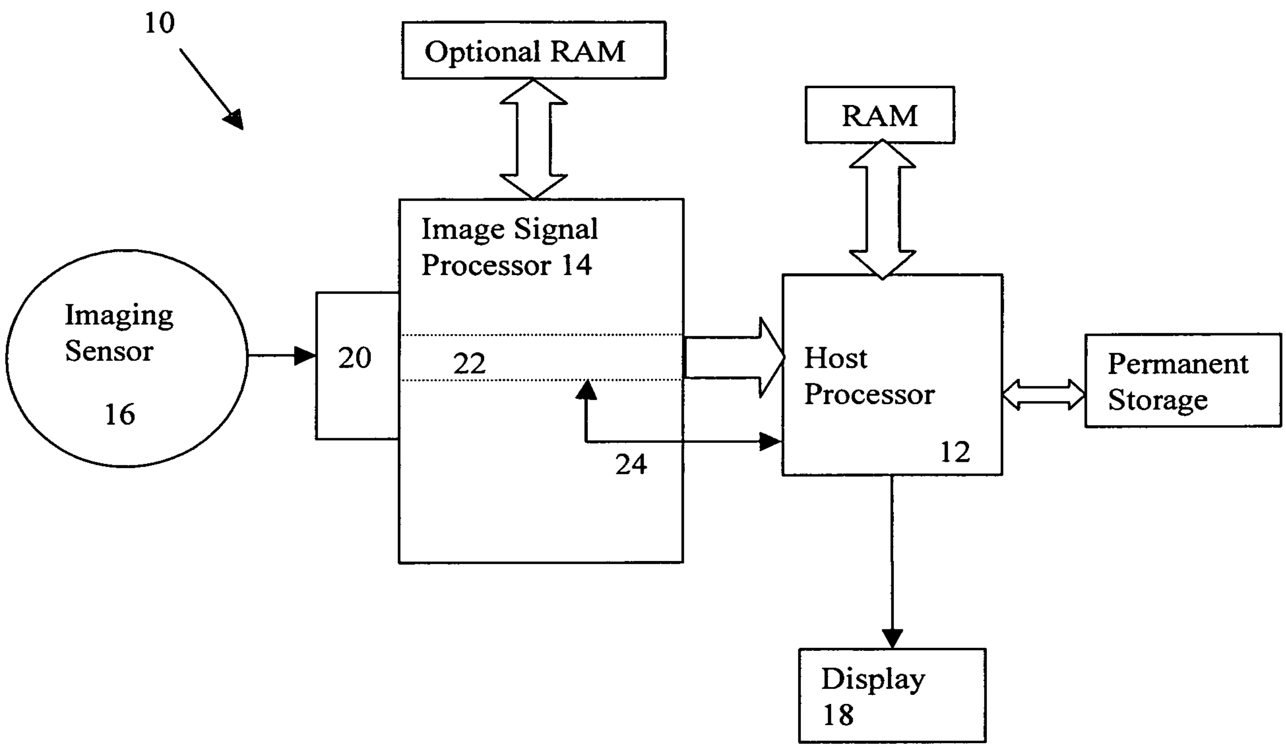

10
Optional RAM
RAM
Image Signal
Processor 14
Imaging
Sensor
16
20
22
24
Host
Processor
12
Permanent
Storage
Display
18

Figure 1

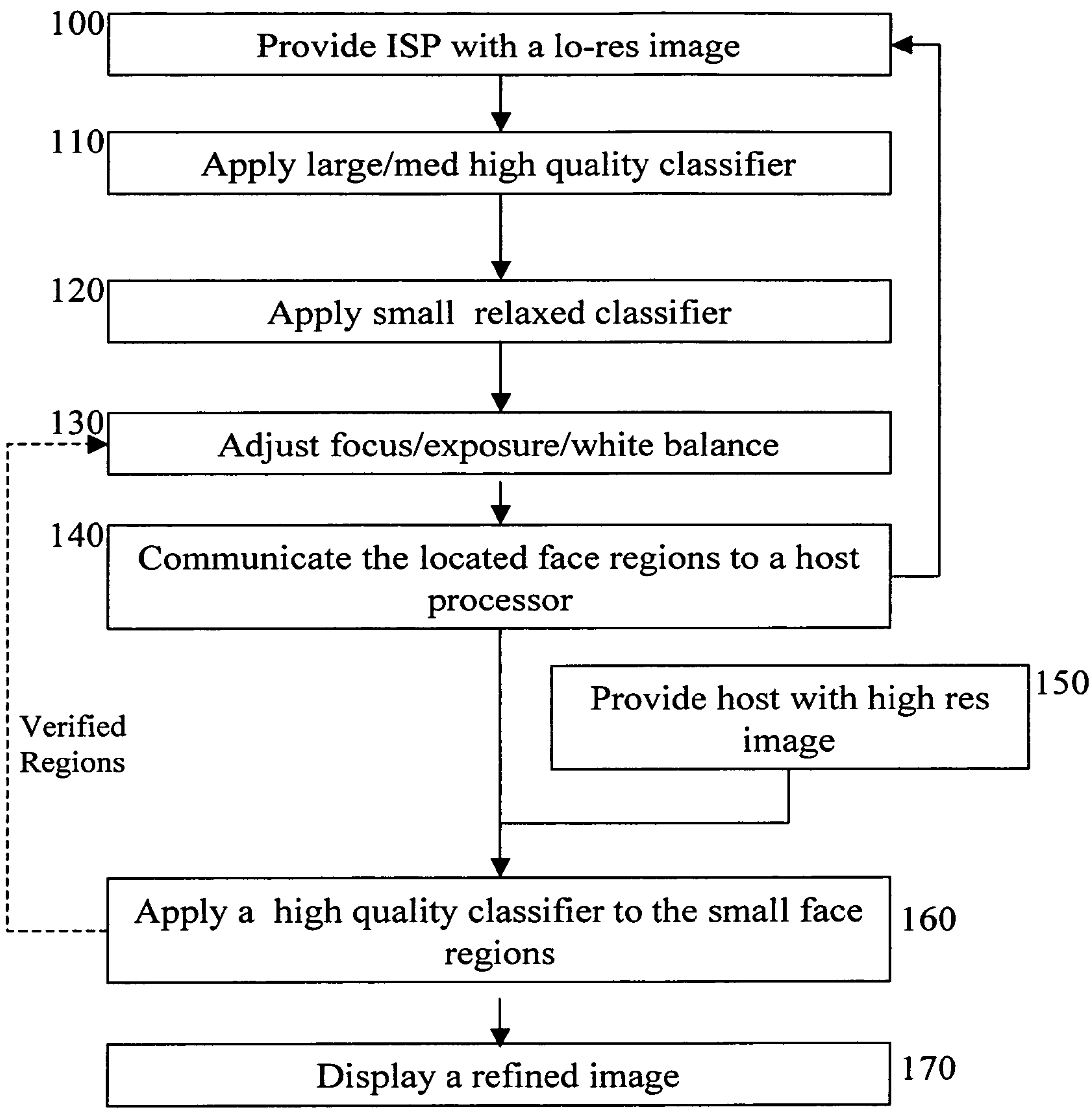

100
Provide ISP with a lo-res image
110
Apply large/med high quality classifier
120
Apply small relaxed classifier
130
Adjust focus/exposure/white balance
140
Communicate the located face regions to a host processor
Provide host with high res image
150
Verified Regions
Apply a high quality classifier to the small face regions
160
Display a refined image
170

Figure 2

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2007/009763

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06K9/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SEBASTIEN ROUX ET AL: "Embedded Convolutional Face Finder" MULTIMEDIA AND EXPO, 2006 IEEE INTERNATIONAL CONFERENCE ON, IEEE, PI, 1 July 2006 (2006-07-01), pages 285-288, XP031032828 ISBN: 978-1-4244-0366-0	1,11,12
Y	the whole document	2-10
Y	EP 1 785 914 A (SONY CORP [JP]) 16 May 2007 (2007-05-16) abstract; figures 4,7a,7b paragraphs [0048], [0080] - [0084], [0088] - [0090]	2-10

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 June 2008	08/07/2008

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Rimassa, Simone

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2007/009763

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MING-HSUAN YANG; NARENDRA AHUJA: "Face detection and gesture recognition for human-computer interaction" 2004, KLUWER ACADEMIC PUBLISHERS , UNITED KINGDOM , XP002484603 page 33 - page 35; figure 2.10	1,11,12
A	HEISELE B ET AL: "Hierarchical classification and feature reduction for fast face detection with support vector machines" PATTERN RECOGNITION, ELSEVIER, GB, vol. 36, no. 9, 1 September 2003 (2003-09-01), pages 2007-2017, XP004429549 ISSN: 0031-3203 section 3. Hierarchy of classifiers abstract	1,11,12
A	US 2007/053614 A1 (MORI KATSUHIKO [JP] ET AL) 8 March 2007 (2007-03-08) abstract; figure 5 paragraph [0053]	1,11,12

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2007/009763

Patent document cited in search report		Publication date	Patent family member(s)			Publication date
EP 1785914	A	16-05-2007	CN	1967566	A	23-05-2007
			JP	2007135115	A	31-05-2007
			KR	20070051689	A	18-05-2007
			US	2007110321	A1	17-05-2007
US 2007053614	A1	08-03-2007	CN	1928887	A	14-03-2007
			JP	2007072606	A	22-03-2007